UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/876,714	06/07/2001	Paul M. Dantzig	YOR9-2001-0261US1 (8728-5	9392

7590 06/29/2004

F. CHAU & ASSOCIATES, LLP
Suite 501
1900 Hempstead Turnpike
East Meadow, NY 11554

EXAMINER	
ROSWELL, MICHAEL	

ART UNIT	PAPER NUMBER
2173	8

DATE MAILED: 06/29/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/876,714	DANTZIG ET AL.	
	Examiner	Art Unit	
	Michael Roswell	2173	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *20 March 2002*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-33* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-33* is/are rejected.

7)☒ Claim(s) *16,19,25 and 26* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *19 September 2001* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date *6*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Claim Objections

Claims 16, 19, 25, and 26 are objected to because of the following informalities: improper dependent claim format. Appropriate correction is required.

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

Claims 1-3, 8, 11-13, and 27-30 are rejected under 35 U.S.C. 102(b) as being anticipated by Raman (U.S. Patent 5,748,186).

Regarding claim 1, Raman teaches retrieving a modality-independent document from local storage (taught as the storing of a "common intermediate high-level data structure" stored in memory, which is then retrieved by a presentor, at col. 3, lines 6-11), parsing the modality-independent document using parsing rules obtained from local or remote storage (taught as the parsing of a source document, which must inherently contain rules for parsing on local or remote storage, at col. 4, lines 45-49), converting the modality-independent document into a first intermediate representation that can be rendered by a speech user interface and converting the modality-independent document into a second intermediate representation that can be rendered by a graphical user interface (taught as the conversion of the common intermediate structure into aural information or visual information, at col. 3, lines 11-13), building a cross-reference

table by which the speech user interface can access components comprising the second intermediate representation (taught as the ability of a user to interact with forms through speech input, at col. 4, lines 21-27. The program must inherently contain speech definitions for accessing the graphical interface), rendering the first and second intermediate representations in their respective modalities (taught as the use of rendering methods to render a document into a specified modality, at col. 7, lines 53-57), and receiving a user input in one of the GUI and speech user interface modalities to enable multi-modal interaction and control the document presentation (taught as the use of an interactive interface to control the document, at col. 3, lines 31-35).

Regarding claim 2, Raman teaches synchronizing GUI and speech modalities at col. 3, lines 21-23.

Regarding claim 3, Raman teaches storing the first intermediate representation in a local system memory for immediate rendering, taught as the storing of a common structure in memory, at col. 3, lines 6-8.

Regarding claim 8, Raman teaches executing an applications program corresponding to an event call within the modality-independent document, taught as the execution of a browser to interact with links specified in the common intermediate structure, at col. 3, lines 47-51, and 56-62.

Regarding claim 11, Raman teaches a method for registering a program to be executed upon completion of a user-specified event, taught as the use of event methods to bind a

presentation method and thus an application to a document object for a particular event, at col. 7, lines 33-38.

Regarding claim 12, Raman teaches a modality-independent document comprising an intent-based document, taught as the use of HTML documents for conversion in to a common intermediate structure, at col. 3, lines 36-46.

Regarding claim 13, the system of Raman is inherently composed of computer-executable instructions and stored on a machine-readable storage device.

Regarding claim 27, Raman teaches a multi-modal manager for parsing a modality-independent document to generate a traversal model that maps components of a modality-independent document to first and second modality-specific representations (taught as the use of a recognizer to convert a common intermediate structure into visual, aural, or tactile information, at col. 3, lines 6-8), a speech user interface manager for rendering and presenting a first modality-specific representation in a speech modality and a GUI manager for rendering and presenting the second modality-specific representation of a GUI modality (taught as the use of a presentor for presenting aural and visual information to a user, at col. 3, lines 8-16), an event queue monitor for detecting GUI events and an event queue for storing captured GUI events (inherently taught as the I/O control of external devices by an operating system), and a plurality of methods, called by a speech user interface manager for synchronizing I/O events across speech and GUI modalities (taught as the concurrently processed navigational methods of col. 5, lines 39-47).

Regarding claim 28, Raman teaches the methods for synchronizing I/O events comprising a first method for polling for the occurrence of GUI events in the event queue and a second method for reflecting speech events back to the GUI manager and posting speech events to the multi-modal manager, taught as the use of visible and audible navigational methods associated with objects at col. 5, lines 39-47, and combined with the inherent I/O queues of an operating system as shown *supra*.

Regarding claim 29, Raman teaches a method for invoking user-specified programs that are specified in the modality-independent document, taught as the use of event methods to bind a presentation method and thus an application to a document object for a particular event, at col. 7, lines 33-38.

Regarding claim 30, Raman teaches a multi-modal manager comprising a main renderer that instantiates a GUI manager, a speech user interface manager, and a method for capturing GUI events, taught as the use of a recognizer for converting a document into a common intermediate structure, which in turn instantiates a presentor for presenting data graphically or aurally, at col. 3, lines 6-16, and an interactive interface controlling I/O devices, at col. 3, lines 30-34.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said

subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 4-7, 9, 10, 14-26, 32, and 33 are rejected under 35 U.S.C. 103(a) as being unpatentable over Raman and Ehsani et al (U.S. Publication 2002/0032564), hereinafter Ehsani.

Regarding claim 4, Raman teaches converting a modality-independent document to a first intermediate representation, taught as the conversion of a common intermediate structure into aural information or visual information, at col. 3, lines 11-13.

Raman fails to explicitly teach transcoding a modality-independent document to a speech markup script.

Ehsani teaches the generation of recognition grammars from source pages to be used in a speech interface similar to that of Raman, at ¶ 0022 of the disclosure. Ehsani also teaches transcoding a modality-independent document to a speech markup script, taught as the implementation of a voice page in Voice XML, at ¶ 0231.

Therefore, it would have been obvious to one of ordinary skill in the art, having the teachings of Raman and Ehsani before him at the time the invention was made to modify the speech interface of Raman to include the transcoding of a modality-independent document into Voice XML of Ehsani in order to obtain a speech user interface with voice scripting capabilities.

One would be motivated to make such a combination for the advantage of allowing a user to interact with an application vocally and defining the vocal interactions in a highly structured language such as Voice XML. See Ehsani, ¶ 0230.

Regarding claim 5, Raman teaches deferred rendering of a speech markup script, taught as the user selection of a presentation modality, at col. 2, lines 49-51.

Regarding claim 6, Ehsani teaches storing the speech markup script on a local persistent storage device, taught as the use of a voice server for storing the voice pages, where the voice server is also the apparatus used to implement the Web-based application, or client computer, at ¶ 0232.

Regarding claim 7, Ehsani teaches creating a speech markup script in VXML, at ¶ 0231.

Regarding claims 9 and 10, Ehsani teaches updating existing grammar rules with data values returned from the applications program and updating content values associated with a component of the modality-independent document using data values returned from the applications program, taught as the editing of grammars at ¶ 0245 and listing of option values at ¶ 0244.

Regarding claim 14, Raman teaches preparing an internal representation of a structure and component attributes of a modality-independent document, taught as the parsing of a source document and creation of an element tree, at col. 5, lines 1-10, and presenting an aural description of the modality-independent document in response to the spoken request, taught as the "speaking" of a document (col. 4, lines 6-7), in response to the speech recognizer control of the presentor module, at col. 3, lines 31-35.

Raman fails to explicitly teach building a grammar comprising rules for resolving specific spoken requests and processing a spoken request utilizing the grammar rules.

Ehsani teaches the generation of recognition grammars from source pages to be used in a speech interface similar to that of Raman, at ¶ 0022 of the disclosure. Furthermore, Ehsani

teaches building a grammar comprising rules for resolving specific spoken requests (taught as the generation of a grammar through a voice page, at ¶ 0237-0246), and processing a spoken request utilizing the grammar rules (taught as the use of a voice recognition system to understand user statements, at ¶ 0248).

Therefore, it would have been obvious to one of ordinary skill in the art, having the teachings of Raman and Ehsani before him at the time the invention was made to modify the multiple-modality presentation system of Raman to include the grammar building system of Raman, in order to obtain a presentation system capable of constructing document-specific vocal grammars.

One would be motivated to make such a combination for the advantage of document portability. See Ehsani, ¶ 0009.

Regarding claim 15, Raman teaches presenting an aural description of a modality-independent document by presenting document components, attributes, or methods of interaction, taught as the aural presentation of document data, which inherently contains components, attributes, and methods, at col. 5, lines 21-47.

Regarding claim 16, the system of Raman is inherently composed of computer-executable instructions and stored on a machine-readable storage device.

Regarding claim 17, Raman teaches preparing an internal representation of a structure and component attributes of a modality-independent document, taught as the parsing of a source document and creation of an element tree, at col. 5, lines 1-10, and presenting an aural description of a modality-independent document by presenting document components,

attributes, or methods of interaction, taught as the aural presentation of document data, which inherently contains components, attributes, and methods, at col. 5, lines 21-47.

Raman fails to explicitly teach building a grammar comprising rules for resolving specific spoken requests and processing a spoken request utilizing the grammar rules.

Ehsani teaches the generation of recognition grammars from source pages to be used in a speech interface similar to that of Raman, at ¶ 0022 of the disclosure. Furthermore, Ehsani teaches building a grammar comprising rules for resolving specific spoken requests (taught as the generation of a grammar through a voice page, at ¶ 0237-0246), and processing a spoken request utilizing the grammar rules (taught as the use of a voice recognition system to understand user statements, at ¶ 0248).

Therefore, it would have been obvious to one of ordinary skill in the art, having the teachings of Raman and Ehsani before him at the time the invention was made to modify the multiple-modality presentation system of Raman to include the grammar building system of Raman, in order to obtain a presentation system capable of constructing document-specific vocal grammars.

One would be motivated to make such a combination for the advantage of document portability. See Ehsani, ¶ 0009.

Regarding claim 18, Ehsani teaches a method wherein building a grammar comprises the step of combining values obtained from data stored local storage or remote storage, with values obtained from an analysis of the modality-independent document, taught as the creation of a grammar based on parsing reference page values to create grammar phrases, at ¶ 0237-0246.

Regarding claim 19, the system of Raman is inherently composed of computer-executable instructions and stored on a machine-readable storage device.

Regarding claim 20, Raman teaches preparing an internal representation of a structure and component attributes of a modality-independent document, taught as the parsing of a source document and creation of an element tree, at col. 5, lines 1-10, and presenting an aural description of the modality-independent document in response to the spoken request, taught as the "speaking" of a document (col. 4, lines 6-7), in response to the speech recognizer control of the presentor module, at col. 3, lines 31-35.

Raman fails to explicitly teach building a grammar comprising rules for resolving specific spoken requests, processing a spoken request utilizing the grammar rules, and obtaining state and value information information regarding specified components of the document from the internal representation of the document.

Ehsani teaches the generation of recognition grammars from source pages to be used in a speech interface similar to that of Raman, at ¶ 0022 of the disclosure. Furthermore, Ehsani teaches building a grammar comprising rules for resolving specific spoken requests (taught as the generation of a grammar through a voice page, at ¶ 0237-0246), and processing a spoken request utilizing the grammar rules (taught as the use of a voice recognition system to understand user statements, at ¶ 0248). Ehsani also teaches obtaining state and value information information regarding specified components of the document from the internal representation of the document, taught as the creation of a grammar based on parsing reference page values to create grammar phrases, at ¶ 0237-0246.

Therefore, it would have been obvious to one of ordinary skill in the art, having the teachings of Raman and Ehsani before him at the time the invention was made to modify the

multiple-modality presentation system of Raman to include the grammar building system of Raman, in order to obtain a presentation system capable of constructing document-specific vocal grammars.

One would be motivated to make such a combination for the advantage of document portability. See Ehsani, ¶ 0009.

Regarding claim 21, Ehsani teaches a method for building a grammar by combining values obtained from data stored in local storage or remote storage, with values obtained from analysis of the document, taught as the creation of a grammar based on parsing reference page values to create grammar phrases, at ¶ 0237-0246.

Regarding claim 22, the system of Raman is inherently composed of computer-executable instructions and stored on a machine-readable storage device.

Regarding claim 23, Raman teaches preparing an internal representation of a structure and component attributes of a modality-independent document, taught as the parsing of a source document and creation of an element tree, at col. 5, lines 1-10, and presenting an aural description of the modality-independent document in response to the spoken request, taught as the "speaking" of a document (col. 4, lines 6-7), in response to the speech recognizer control of the presentor module, at col. 3, lines 31-35. Raman also teaches presenting each character of content value information requested in response to a spoken request, taught as the display of visual information (col. 3, lines 11-16) in response to an I/O manipulation of a presentor by a speech recognizer (col. 3, lines 31-35).

Raman fails to explicitly teach building a grammar comprising rules for resolving specific spoken requests, processing a spoken request utilizing the grammar rules, and obtaining state and value information information regarding specified components of the document from the internal representation of the document.

Ehsani teaches the generation of recognition grammars from source pages to be used in a speech interface similar to that of Raman, at ¶ 0022 of the disclosure. Furthermore, Ehsani teaches building a grammar comprising rules for resolving specific spoken requests (taught as the generation of a grammar through a voice page, at ¶ 0237-0246), and processing a spoken request utilizing the grammar rules (taught as the use of a voice recognition system to understand user statements, at ¶ 0248). Ehsani also teaches obtaining state and value information information regarding specified components of the document from the internal representation of the document, taught as the creation of a grammar based on parsing reference page values to create grammar phrases, at ¶ 0237-0246.

Therefore, it would have been obvious to one of ordinary skill in the art, having the teachings of Raman and Ehsani before him at the time the invention was made to modify the multiple-modality presentation system of Raman to include the grammar building system of Raman, in order to obtain a presentation system capable of constructing document-specific vocal grammars.

One would be motivated to make such a combination for the advantage of document portability. See Ehsani, ¶ 0009.

Regarding claim 24, Raman teaches inserting pauses between each character of the content value information to be presented, taught as the pausing of the presentation on links to enable easier user selection, at col. 4, lines 12-14.

Regarding claim 25, Ehsani teaches a method for building a grammar by combining values obtained from data stored in local storage or remote storage, with values obtained from analysis of the document, taught as the creation of a grammar based on parsing reference page values to create grammar phrases, at ¶ 0237-0246.

Regarding claim 26, the system of Raman is inherently composed of computer-executable instructions and stored on a machine-readable storage device.

Regarding claims 32 and 33, Ehsani teaches the use of Voice XML in creating voice pages from source pages, and allows for the presentation of such pages, at ¶ 0232.

Claim 31 is rejected under 35 U.S.C. 103(a) as being unpatentable over Raman and Dietz (U.S. Patent 6,175,820).

Raman has been shown to teach a speech user interface manager at col. 3, lines 8-16.

However, Raman fails to explicitly teach the use of JSAPI in the speech user interface manager.

Dietz teaches a system for enhanced computerized speech communication, which utilizes JSAPI to interface with the user (col. 1-2, lines 65-67 and 1-2.

Therefore, it would have been obvious to one of ordinary skill in the art, having the teachings of Raman and Dietz before him at the time the invention was made to modify the speech user interface of Raman with the JSAPI capabilities of Dietz in order to obtain a speech user interface with JSAPI speech synthesizers.

One would be motivated to make such a combination for the advantage of the synthesizer output control given by JSAPI and the related Java Speech Markup Language (col. 2, lines 3-9).

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure, and describes the state of the art in general as related to the application.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Michael Roswell whose telephone number is (703) 305-5914. The examiner can normally be reached on 8:30 - 6:00 M-F.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Cabeca can be reached on (703) 308-3116. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Michael Roswell
6/23/2004

CAO (KEVIN) NGUYEN
PRIMARY EXAMINER